1-1-2004

Proline isomerization as a cell signaling control mechanism: Measuring the effect of a molecular switch on interaction behavior in the Itk SH2 domain

Patrick John Breheny
Iowa State University

Proline isomerization as a cell signaling control mechanism: Measuring the effect of a molecular switch on interaction behavior in the Itk SH2 domain

by

Patrick John Breheny

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Biochemistry

Program of Study Committee:
Amy Andreotti, Major Professor
Mark Hargrove
Kai-Ming Ho

Iowa State University

Ames, Iowa

2004

Graduate College

Iowa State University

This is to certify that the master's thesis of

Patrick John Breheny

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

TABLE OF CONTENTS

ABSTRACT

The molecular regulation of protein-protein interactions is often accomplished via discrete structural changes within binding sites. The Src homology 2 (SH2) domain of interleukin-2 tyrosine kinase (Itk) binds two separate ligands *in vitro*: a phosphotyrosine-containing peptide and the Itk SH3 domain. Binding specificity for these two ligands is affected by cis/trans isomerization of the Asn 286-Pro 287 imide bond in the Itk SH2 domain. In this thesis, I describe the development of a novel method of analyzing nuclear magnetic resonance data, in which chemical shift perturbations and cross-peak intensities are correlated in order to measure the independent ligand-binding affinities of each SH2 conformer. This method reveals the cis imide bond-containing Itk SH2 conformer to exhibit a 3.7-fold greater affinity for the Itk SH3 domain compared with the binding of the trans conformer to the same ligand. Conversely, the trans conformer binds the phosphotyrosine-containing peptide with a 3.5-fold greater affinity than the cis conformer. In addition to furthering the understanding of this system, the method presented here is of general application in quantitatively determining the specificities of conformationally heterogeneous systems that use a molecular switch to regulate binding to multiple distinct ligands.

CHAPTER 1: INTRODUCTION AND BACKGROUND

Background and Significance

Kinases, modular domains, and their roles in cellular signaling

Organisms rely on cellular signaling in order to regulate processes in many types of tissues. Great progress has been made in the last two decades to understand the molecular basis by which cells are capable of transmitting messages, and the mechanisms by which the accuracy of these communications is ensured. The paradigm that has emerged is that modular domains – relatively small sequences of around 100 amino acids that adopt specific three-dimensional structures within proteins – play a large role in governing the precision of cellular signaling pathways. These domains influence cellular signaling not through direct catalysis, but rather through protein-protein interactions. Understanding the structural mechanisms by which protein-protein interactions occur and are controlled has thus become a central goal not only in the understanding of cellular communication, but of biochemistry in general.

The first protein in which these domains were observed and their importance realized was the tyrosine kinase Src. Kinases are an extremely important group of enzymes responsible for the transfer of phosphoryl groups, usually from ATP to a specific molecular target. This transfer is a fundamental molecular switch and plays a central role in the regulation of numerous cellular pathways. Eukaryotic kinases share a fair amount of sequence homology, and are composed of two subgroups: those that phosphorylate serine or threonine, and those that phosphorylate tyrosine. Due to their vital cellular roles, including the regulation of mitosis, many oncogenes code for tyrosine kinases with defective regulatory mechanisms[1], thereby underscoring the importance of proper control of these kinases.

The study of Src revealed that, in addition to the kinase domain, the protein contains two noncatalytic elements with sequence homology to elements found in other proteins[2,3]. These elements, termed SH2 and SH3 (Src Homology 2 and 3) domains were discovered to have a folded structure[2] and function[4] independent of the kinase domain, yet were essential for the proper localization of Src and strongly affected kinase activity. These domains have since been discovered in hundreds of proteins, and many other types of domains have been discovered since these initial landmark discoveries, including PH, PTB, PDZ, and WW domains[5]. Protein modules have now been observed to play fairly universally conserved roles in a wide variety of cellular processes.

The discovery and characterization of these protein modules has revolutionized our understanding of the organization and evolution of cell signaling pathways. With the emerging model of the living cell as a highly structured and organized entity, the importance of enzyme targeting and localization within the cell is being increasingly realized, as is the role of protein-protein interaction domains in organizing the assembly of these protein networks[6].

The past decade has also seen the determination of the crystal structures of a wide variety of protein kinases[7]. These structures have suggested that the catalytic domains of eukaryotic kinases, in addition to their sequence homology, are also structurally conserved. However, these same crystallographic analyses revealed that despite the similarity of kinase domain structure, there exists a wide variety of means employed by the cell to control the activation and deactivation of these kinases. Understanding the mechanisms that underlie the regulation of kinases has therefore proven to be an important area of continuing research.

Further structural studies of Src and the family of proteins to which it belongs (the Src-family) revealed a particularly fascinating system of control, in which its SH2 and SH3 domains play a direct role in allosteric regulation of the kinase domain[8]. To date, Src-family kinases are the only nonreceptor tyrosine kinases for which the mechanism of activation and

deactivation of catalytic activity is well understood. The Src-family has thus become a model for the manner in which protein domain-mediated tyrosine kinase regulation is achieved.

Insights into nonreceptor tyrosine kinase regulation from the Src-family

The Src family is the largest family of nonreceptor tyrosine kinases. All Src family members share a common domain organization that provides the foundation for what seems to be a conserved regulatory mechanism (Figure 1-1). Each protein contains an N-terminal region responsible for membrane localization (called an SH4 domain), followed by a variable region of 40-70 residues. Continuing in the C-direction, Src kinases then contain an SH3 domain, an SH2 domain, a kinase domain, and a C-terminal tail[8].

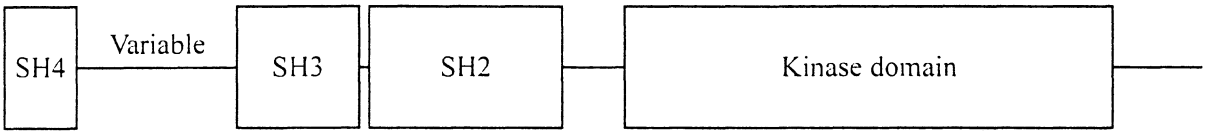


Figure 1-1. Domain structure of the Src family kinases. Only the SH3 through C-terminus fragment is required for the inhibition of the kinase domain, which is catalytically active when expressed alone.

The mechanism of Src regulation emerged following the determination of several key structures of Src-family protein fragments, all within a short window of time. First, a Src-family kinase domain (Lck) was crystallized in the active form[9]. Soon after, the structures of two family member fragments including their regulatory domains (Src and Hck), were determined in the down-regulated form[10,11]. Together, this set of structures has suggested a general model for the regulation of Src-family kinases.

The structure of the catalytic domain of Lck in the active state is very similar to other structures of active kinase domains, most notably that of protein kinase A[12]. Like other kinase domains, Src-family kinase domains consist of two lobes: a smaller N-terminal lobe, and a larger C-terminal lobe. The ATP-binding site and catalytic cleft are located in the interface between the lobes.

Unexpectedly, however, the crystallization of the larger fragments of Src and Hck revealed a novel contribution to the kinase structure provided by the SH3 and SH2 domains, as well as the C-terminal tail. The SH2 domain was observed bound to a phosphotyrosine residue in the tail, and the SH3 domain was seen bound to a polyproline helix in the linker between the SH2 domain and the kinase. The effect of these interactions, which occur on the opposite side of the kinase domain from the catalytic cleft, is to pull on each lobe, displacing key structural features of the catalytic cleft. These interesting structural findings reveal that, in addition to their role in targeting, SH2 and SH3 domains are capable of regulating the kinase domains to which they are attached, a finding which may turn out to be a common feature of tyrosine kinase control.

The cellular role(s) of Itk

The significant advances that have been made in understanding signaling pathways in the past two decades have also shed considerable light on the process of the engagement of immune response. Tyrosine kinases, particularly those of the Tec family, have been demonstrated to play critical roles in this process. Among them, the Tec kinase Itk (Emt/Tsk) plays a large role in modulating T cell signaling pathways[13].

The experiments that have been conducted thus far have suggested the following general model for Itk's role in T cell signal transduction following engagement of the antigen receptor. The phosphorylation of LAT (linker for activation of T cells) by Zap70 leads to the recruitment of multiple signaling molecules, including Itk[14,15]. Once there, it is

phosphorylated by Lck on Tyr 511 in the activation loop of the kinase domain, which leads to the activation of Itk[16]. Itk activation in turn leads to the phosphorylation and activation of PLC-γ1, a crucial event in calcium mobilization that in turn leads to T cell activation[17].

While there is ample experimental evidence describing the biology of Itk during T cell activation, much less is known about the structural differences between active and inactive Itk. The crystal structures of the Itk kinase domain in phosphorylated and dephosphorylated states reveal that only minor structural differences occur in the kinase domain upon phosphorylation of Tyr 511, and that additional effects contributed by the regulatory domains appear to be necessary in order to explain the kinase activation[18].

This, of course, would be similar to the case in the Src family, where the SH2 and SH3 domains are necessary to modify the structure of the kinase domain. Like the Src kinases, Itk contains a kinase domain preceded by SH2 and SH3 domains. However, unlike Src kinases, Itk and the other Tec kinases do not contain the C-terminal tail so crucial to Src regulation, which indicates that the activation of the kinase domain must occur via a different mechanism, also seemingly involving the regulatory domains.

The domain structure of Itk, in addition to the features mentioned above, includes an N-terminal pleckstrin homology (PH) domain that is responsible for the recruitment of Itk to the cell membrane, and a Tec homology (TH) domain, which includes a proline-rich region immediately upstream of the SH3 domain. The domain structure of Itk is presented in Figure 1-2, which also discusses some of the protein-protein interactions that each domain has been implicated in. As can be seen by the number of different interactions they take part in, the regulatory domains may have multiple roles depending on the environment and regulatory state of Itk.

The structures of the Itk kinase domain are the only crystal structures available for the protein. In the absence of data from X-ray crystallography regarding the structures of active and inactive Itk, information regarding the mechanism of kinase regulation have come from

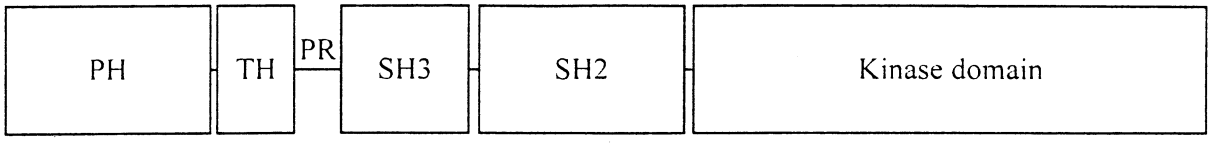


Figure 1-2. Domain structure of Itk. The domains and their known roles are as follows[19]. PH (pleckstrin homology) domains interact with phospholipids as well as a variety of proteins usually found near the cell surface. The Itk PH domain is believed responsible for recruitment of Itk to the cell membrane signaling complexes. The function of TH (Tec homology) domains are not yet clear, although the PR (proline-rich) region is an SH3 ligand and forms an intramolecular complex with the neighboring SH3 domain. SH3 domains bind proline-rich regions, and in Itk, also interact in an intermolecular fashion with the Itk SH2 domain. The SH2 domain, in addition its SH3 interaction, also binds phosphorylated tyrosine residues in certain proteins. All Tec kinases possess SH3, SH2, and kinase domains, although the domain organization at the N-terminus is somewhat variable between its members. The above organization, however, is shared by Itk, Tec, and Btk.

NMR spectroscopy. In recent years, the structures of the Itk SH2 domain and the Itk PR-SH3 fragment have been solved using NMR[20,21]. The structure of the PR-SH3 fragment revealed an intramolecular interaction between the proline rich region and the SH3 domain, while another NMR study has demonstrated that the SH3-SH2 fragment of Itk dimerizes[22]. While the functional significance of these structural findings is unclear, these inter- and intramolecular regulatory domain interactions provide tantalizing clues suggesting the possible mechanisms that may be at play in Itk activation and deactivation.

The SH2 and SH3 domains of Itk may function to provide a mechanism for the direct regulation of the kinase, as they do in Src. Alternatively, their principal function may be to affect cellular localization. Their role could also involve other possibilities, such as the binding of an inhibitor or effector, or indeed some combination of the above roles. Regardless, however, of the role they play, these domains are certainly important for Itk cellular function and the means by which these roles are structurally modulated is an interesting question that is currently best answered using NMR spectroscopy.

Approach

NMR is a powerful tool to probe regulatory domain interactions

Nuclear magnetic resonance (NMR), apart from its utility as a structure-determination method, has a wide range of applications to the study of protein-protein interactions. NMR techniques have been developed to map binding surfaces, reveal intermolecular contacts between atoms, investigate allosteric effects, study induced changes in protein dynamics, and measure the affinity, specificity, and kinetics of binding[23-25]. Due to the wealth of information obtained by NMR, there is a large and growing body of data collection and analysis methods available with which to study molecular interactions.

The tractable nature of regulatory domains makes them ideal candidates for study by NMR. Regulatory domains can be easily and relatively inexpensively labeled with ^{15}N by growing an *E. coli* production strain in minimal media using $^{15}NH_4Cl$ as the sole nitrogen source. Once grown, these proteins can be harvested, purified, and studied in a variety of ways by NMR.

HSQC spectroscopy

The most common NMR experiment carried out with ^{15}N-labeled protein is the ^{1}H-^{15}N heteronuclear single-quantum correlation experiment, or HSQC. This experiment yields a two-dimensional spectrum correlating the chemical shifts of protons with the chemical shifts of attached, isotopically-labeled, nitrogens. In an ^{15}N-labeled protein, then, signals are seen for all protons that are covalently attached to nitrogen. The spectrum contains peaks for the side-chains of tryptophan, lysine, and arginine, but the most common signal is given by backbone amide groups. Excepting proline (which has no amide protons), the backbone of every amino acid therefore produces a single peak and serves as an excellent reporter for the chemical environment of the residue. An HSQC spectrum is shown in Figure 2-1. A detailed description (including pulse diagrams) of the HSQC experiment can be found in Ref. 24.

Because the HSQC spectrum contains roughly one signal for each amino acid, the experiment is often said to provide a "fingerprint" of the protein. The most common usage of this fingerprint for the study of protein-protein interactions is the mapping of chemical shift perturbations. In perturbation mapping, the HSQC of one protein is monitored while an unlabeled binding partner is added. An interaction between the proteins causes changes in the local chemical environment of each residue on the binding surface. Due to the high sensitivity of chemical shifts, these changes are easily observed in the spectra. Residues for

which the magnitude of chemical shift is significant can then be mapped onto the surface representation of a protein in order to delineate a contiguous binding surface.

NMR titrations

The study of chemical shift perturbation can be done in a quantitative manner as well. The analysis of NMR titrations varies depending on the exchange regime of the interaction event. Exchange regimes, and issues that arise in the quantitation of NMR data, are discussed below.

Exchange regimes

Chemical exchange between NMR peaks of different frequencies can manifest itself in one of three ways, depending on the exchange rate of the event relative to the NMR timescale: *slow exchange*, in which a peak at both frequencies can be observed and monitored independently; *fast exchange*, in which a single peak is observed, whose position corresponds to the weighted average of the two individual chemical shifts; and *intermediate exchange*, in which the signals become poorly defined, extensive broadening occurs, and the peaks may become indistinguishable from noise. Unlike fast and slow exchange, the poorly resolved data given by intermediate exchange generally precludes quantitative analysis in an NMR titration.

Concentration measurements of the protein substates giving rise to the spectral signals differ depending on the exchange regime. Fast exchange events give rise to signals that vary continuously between the chemical shifts of the free and bound protein states, depending on the concentration of each state. This type of exchange is quantified, then, by following the chemical shift of the center of a resonance over the course of a titration. Fast exchange is typically observed for fairly weak protein-protein interactions. Figure 1-3 demonstrates the way in which fast exchange appears in an NMR spectrum.

Stronger binding events, along with slow/irreversible modifications, such as phosphorylation, give rise to slow exchange. In order to quantify slow exchange, one measures the intensities of the signals given off by each of the two states, thereby obtaining their population ratio. While this sounds trivial, in practice there are many concerns associated with obtaining reliable quantitative data from these measurements.

Measurement of NMR signal intensity

The intensity of an NMR peak is most accurately measured by integrating the three-dimensional volume of the NMR peak, as opposed to simply observing the height of the peak at maximum amplitude. There are two reasons for this: (1) spectral noise will average to zero over the area of integration, and (2) the volume of an NMR peak is independent of broadening, which can occur in an NMR spectrum for a number of reasons, including T_2 relaxation and ligand binding.

In protein NMR spectroscopy, the intensity of a signal is affected by a number of factors (e.g. coupling constants, NMR field strength, buffer conditions). When comparing the signal intensity given off by the same residue in separate states of the same protein in the same NMR tube, however, almost all of these factors are equivalent between the two states. The signal intensity in this case is linearly proportional to the number of protons in the sample giving rise to the signal (i.e. the population of the state). One possible source of discrepancy, however, is a difference between the two states in the ^1H T_1 relaxation rate for a given amide group to be measured in the titration. This rate depends on the internal motions and local flexibility of the protein, and governs the rate at which magnetization of that proton is returned to its equilibrium magnitude. If a proton in one of the two states does not recover the same amount of magnetization between pulses as a proton in the other, the two protons will not give the same intensity of signal, and thereby skew the measurement of the

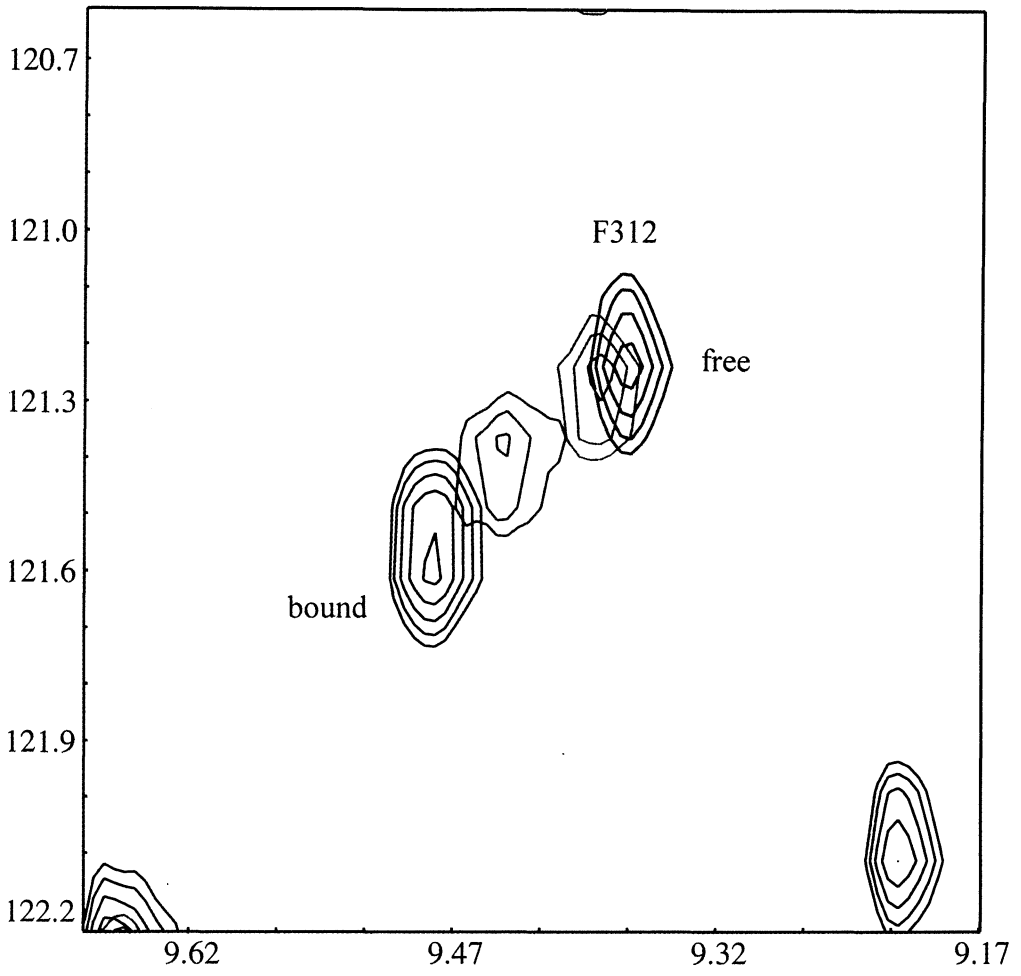


FIGURE 1-3. Select region of the HSQC spectrum for 15N-labeled Itk SH2 upon addition of a phosphopeptide ligand (black – 0.32 mM Itk SH2; green – 0.31 SH2 + 0.06 mM pY; red – 0.31 SH2 + 0.17 mM pY; blue – 0.29 SH2 + 2.58 mM pY). The black peak is the position of the HSQC peak for Phe 312 in the absence of ligand. The blue peak represents SH2 after the addition of 9-fold excess of phosphopeptide, which is sufficient to bind >99% of the SH2 in solution, and therefore depicts Phe 312 in the pY-bound state. This figure is an illustration of an exchange process that is faster than the NMR timescale. For a depiction of slow exchange, see Figure 2-4.

population ratio of the two states. These effects, while ignored in most studies, can nevertheless be significant in certain protein regions.

In addition to requiring relaxation uniformity between substates, the residues used for a quantitative analysis of slow exchange must also be chosen for their isolation from other peaks so as to guarantee the ability to include adequate integration boundaries without errors due to peak overlap. However, if these two criteria are met, integrating peak volumes will correctly correspond to the cis/trans ratio provided that the peaks are accurately integrated.

Thesis Organization

In this chapter, I have given a general introduction to the background of signal transduction kinases and some details of NMR spectroscopy that have proved relevant in conducting the research outlined in the rest of this thesis. Chapter 2 describes the development of a method for analyzing NMR data in order to quantify the independent ligand-binding affinities of each conformer in a conformationally heterogenous protein, the Itk SH2 domain. Chapter 3 describes the future directions of this work, presents preliminary data to highlight those directions, and discusses the significance of this research.

References

1. Martin G. S. Rous sarcoma virus: a function required for the maintenance of the transformed state. *Nature* **227**, 1021-1023 (1970).
2. Sadowski I., Stone J. C., & Pawson T. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus $P130^{gag\text{-}fps}$. *Mol. Cell. Biol.* **6**, 4396-4408 (1986).

3. Mayer B. J., Hamaguchi M., & Hanafusa H. A novel viral oncogene with structural similarity to phospholipase C. *Nature* **332**, 272-275 (1988).
4. Moran M. F., Koch C. A., Anderson D., Ellis C., England L., Martin G. S., & Pawson T. Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proc. Natl. Acad. Sci.* **87**, 8622-8626 (1990).
5. Sudol M. From Src homology domains to other signaling modules: proposal of the 'protein recognition code'. *Oncogene* **17**, 1469-1474 (1998).
6. Pawson T. & Nash P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445-452 (2003).
7. Johnson L. N., Noble M. E. M., & Owen D. J. Active and inactive protein kinases: structural basis for regulation. *Cell* **85**, 149-158 (1996).
8. Sicheri F. & Kuriyan J. Structures of Src-family tyrosine kinases. *Curr Op. Struct. Biol.* **7**, 777-785 (1997).
9. Yamaguchi H. & Hendrickson W. A. Structural bases for activation of the human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature* **384**, 484-489 (1996).
10. Xu W., Harrison S. C., & Eck M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595-602 (1997).
11. Sicheri F., Moarefi I., & Kuriyan J. Crystal structure of the Src-family tyrosine kinase Hck. *Nature* **385**, 602-609 (1997).
12. Zheng J., Trafny E. A., Knighton D. R., Xuong N.-H., Taylor S. S., Ten Eyck L. F., & Sowadski J. M. 2.2 Å refined crystal structure of the catalytic subunit of camp-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta. Crystallogr. D.* **49**, 362-365 (1993).
13. Miller A. T. & Berg L. J. New insights into the regulation and functions of Tec family tyrosine kinases in the immune system. *Curr. Op. Immunol.* **14**, 331-340 (2002).

14. Shan X. & Wange R. L. Itk/Emt/Tsk activation in response to CD3 cross-linking in Jurkat T cells requires ZAP-70 and Lat and is independent of membrane recruitment. *J. Biol. Chem.* **274**, 29323-29330 (1999).

15. Zhang W., Sloan-Lancaster J., Kitchen J., Trible R. P., & Samelson L. E. LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. *Cell* **92**, 83-92 (1998).

16. Heyeck S. D., Wilcox H. M., Bunnell S. C., & Berg L. J. Lck phosphorylates the activation look tyrosine of the Itk kinase domain and activates Itk kinase activity. *J. Biol. Chem.* **272**, 25401-25408 (1997).

17. Bunnell S. C., Diehn M., Yaffe M. B., Findell P. R., Cantley L. C., & Berg L. J. Biochemical interactions integrating Itk with the T cell receptor-initiated signaling cascade. *J. Biol. Chem.* **275**, 2219-30 (2000).

18. Brown K., Long J. M., Vial S. C., Dedi N., Dunster N. J., Renwick S. B., Tanner A. J., Frantz J. D., Fleming M. A., & Cheetham G. M. Crystal structures of interleukin-2 tyrosine kinase and their implications for design of selective inhibitors. *J. Biol. Chem.* **279**, 18727-18732 (2004).

19. Takesono A., Finkelstein L. D., & Schwartzberg P.L. Beyond calcium: new signaling pathways for Tec family kinases. *J. Cell Sci.* **115**, 3039-3048 (2002).

20. Mallis R. J., Brazin K. N, Fulton D. B., & Andreotti A. H. Structural characterization of a proline-driven conformational switch within the Itk SH2 domain. *Nat. Struct. Biol.* **9**, 900-905 (2002).

21. Andreotti A. H., Bunnell S. C., Feng S., Berg L. J., & Schreiber S. L. Regulatory intramolecular association in a tyrosine kinase of the Tec family. *Nature* **385**, 93-97 (1997).

22. Brazin KN, Fulton DB, Andreotti AH. A specific intermolecular association between the regulatory domains of a Tec family kinase. *J. Mol. Biol.* **302**, 607-623 (2000).

23. Zuiderweg E. R. P. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1-7 (2002).

24. Cavanagh J., Fairbrother W. J., Palmer A. G., & Skelton N. J. *Protein NMR Spectroscopy: Principles and Practice*. Academic Press, San Diego (1996).

25. Wuthrich K. *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, Inc., New York (1986).

CHAPTER 2: MODULATION OF LIGAND SPECIFICITY BY PROLYL CIS/TRANS ISOMERIZATION IN THE ITK SH2 DOMAIN

Introduction

Previous *in vitro* investigations studying the SH2 domain of Itk have revealed two intriguing facets of its structure and function: (a) Itk SH2 adopts two stable conformations that are capable of slowly interconverting, and (b) Itk SH2 is capable of independently binding two distinct ligands. The relationship between these two findings and the development of a method to probe that relationship is the central theme of this chapter.

Itk SH2 conformational heterogeneity

Cellular control over a protein's function is often achieved through structural rearrangement, which can be triggered by a variety of mechanisms. Most of the currently known mechanisms involve covalent modification, due to the fact that these modifications are easily and inexpensively detected. Recently, however, it is becoming clear that the structures of a number of proteins are modified in a noncovalent manner[1,2]. The observation of these noncovalent structural modifications is unlikely using most biochemical techniques. NMR, however, is unique in its ability to observe dynamic structural processes, and is thus an incredibly useful technique in the study of such events.

Early NMR spectra of the Itk SH2 domain revealed that the domain adopts two distinct low-energy conformations in solution (Figure 2-1). Varying the temperature at which these spectra were recorded demonstrated that the two conformers freely interconverted in solution, although the slow rate of the conformational exchange suggested the existence of a significant energy barrier between the conformers. This led to the idea that

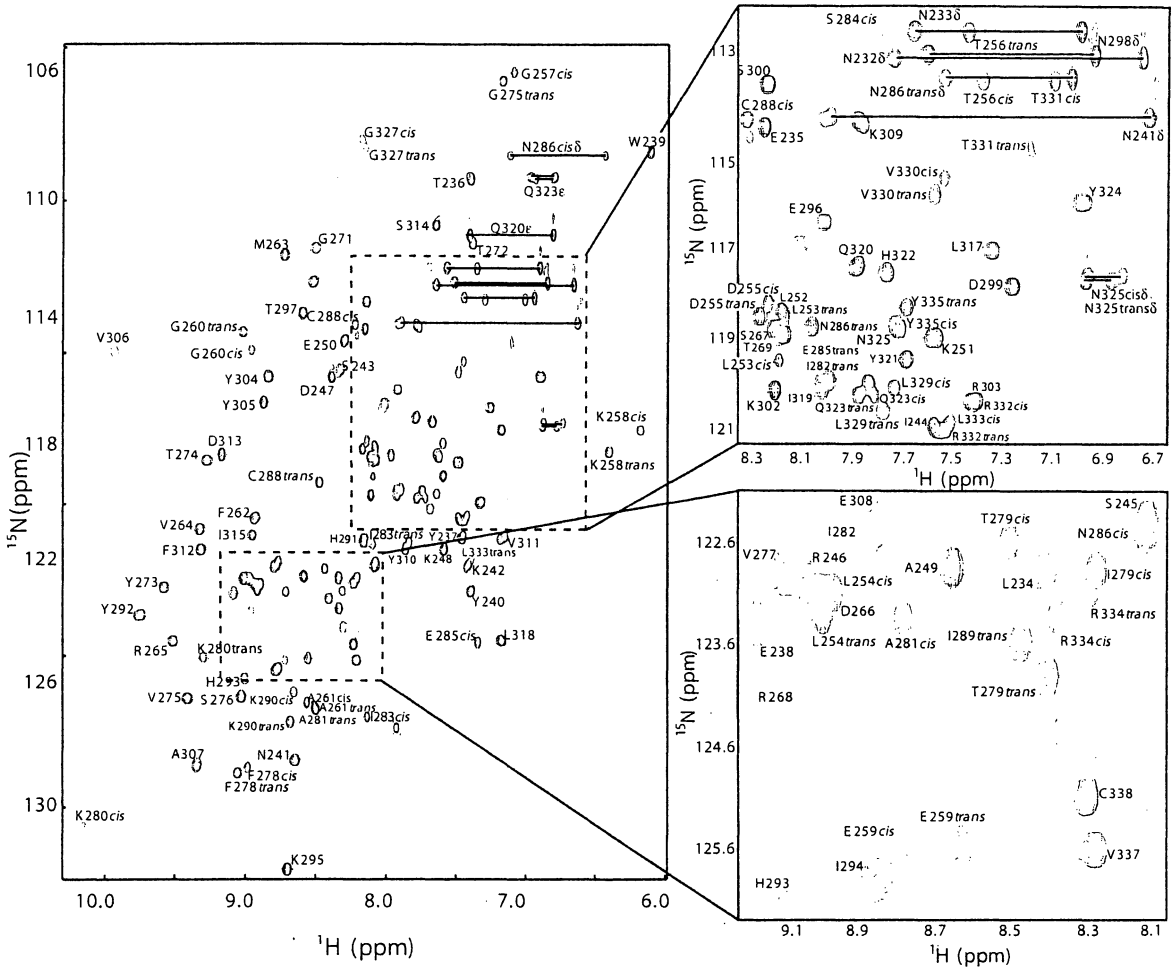


Figure 2-1. ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled Itk SH2. This spectra reveals substantially more backbone amide signals than there are backbone amide groups in the construct. Subsequent NMR resonance assignments showed that this observation was due to conformationally distinct resonance frequencies that occur in 35 of the 109 SH2 residues. Additional NMR experiments demonstrated that the two conformations correlated with the configuration of the Asn 286-Pro 287 imide bond. Regions of high peak density have been expanded for clarity. Figure adapted from R. Mallis et. al.[3]

the barrier was provided by the cis/trans isomerization of a backbone peptidyl-prolyl imide bond. While this isomerization event is fairly common in short peptides, examples for which proline isomerization has been detected in folded proteins is rare, although this is perhaps due to the difficulty of observing isomerization using techniques other than NMR. Nevertheless, characteristic NOE correlations[4] between Asn-286 and Pro-287 verified that the orientation of the imide bond between these residues correlates to each of the two conformations (hereafter, the cis and trans imide bond-containing conformations are referred to as the cis and trans conformations of Itk SH2).

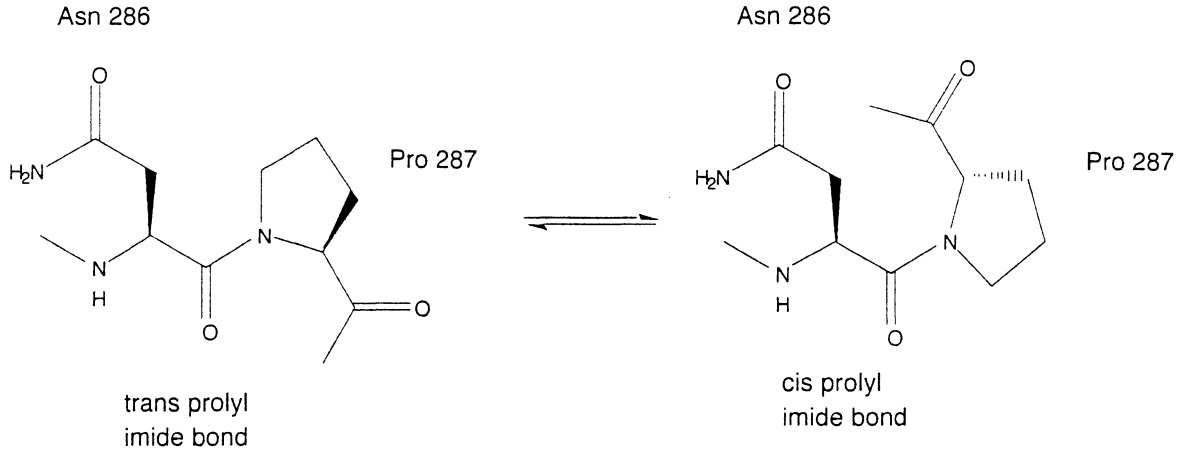


Figure 2-2. Chemical illustration of the prolyl imide bond isomerization event that occurs in Itk SH2.

Among the naturally occurring amino acids, proline is the only residue for which both bond orientations are thermodynamically feasible. A mutation of Pro-287 to glycine, whose bond to the previous residue exclusively adopts the trans orientation, was therefore predicted to populate only the trans conformation. This prediction was then confirmed[5], demonstrating that the conformational heterogeneity of Itk SH2 is dependent on the thermodynamic stability of both orientations of the 286-287 imide bond.

These observations are further supported by the structure determination of both cis and trans conformations by NMR[3], which reveals the significant structural changes in the SH2 domain that occur dependent on the conformation of the imide bond. Surface representations of the Itk SH2 domain illustrating the two conformations and the regions affected by ligand binding are presented in Figure 2-3.

Itk SH2 ligand binding

The Itk SH2 domain is capable of independently binding *in vitro* to two distinct ligands: phosphotyrosine-containing peptides and the Itk SH3 domain[6]. Phosphopeptide binding is common to all SH2 domains, whereas the interaction with an SH3 domain seems, so far, to be unique to Itk.

The structural changes in Itk SH2 that occur as a result of proline isomerization impart different ligand-binding affinities to each conformer. This conformer-specific nature of Itk SH2 ligand recognition is evident in NMR spectra of the protein/ligand complexes. Addition of phosphopeptide to the SH2 domain shifts the equilibrium to favor the trans conformer, resulting in changes in the volumes of the NMR peaks corresponding to each conformer (Figure 2-4 a and b). Also, ligand-induced chemical shift changes are larger for peaks corresponding to residues in the trans conformation than for those corresponding to residues in the cis conformation. In contrast, binding of the Itk SH3 domain to the SH2 domain shifts the cis/trans equilibrium to favor the cis conformer (Figure 2-4 c). In this case, shifts in the positions of HSQC peaks corresponding to the cis conformation, and not that of the trans, are observed in the HSQC spectrum to which recombinant Itk SH3 domain has been added[5]. This seems to be the first demonstration of ligand recognition that is governed by the conformation of a single prolyl imide bond within a folded protein.

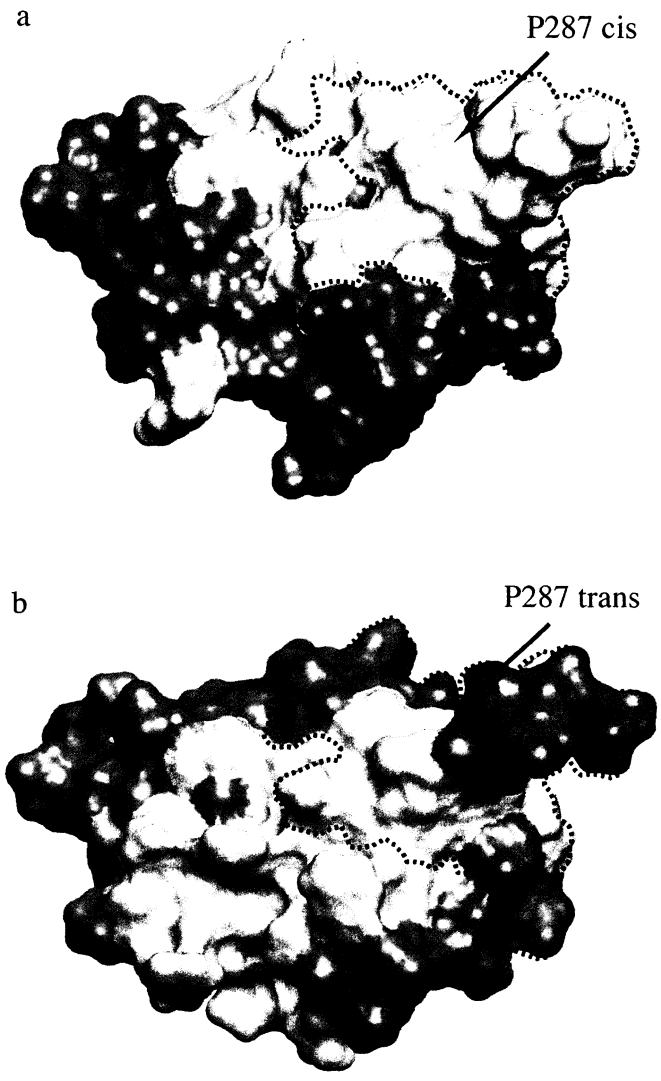


Figure 2-3. Surface representations of the Itk SH2 domain in the (a) cis and (b) trans conformations. On each model, the position of Pro 287 is indicated, and the conformation-doubled residues are demarcated with a dotted line. Chemical shift perturbations that occur on addition of equimolar Itk SH3 domain are mapped onto the cis model and highlighted in gray (a). Similarly, SH2 residues that undergo chemical shift perturbations upon addition of 3-fold excess of phosphopeptide are highlighted in yellow on the trans model (b). For both models, residues that remain unperturbed upon addition of ligand are colored red. Figure rendered with MOLMOL[7]. Adapted from Brazin et. al.[5]

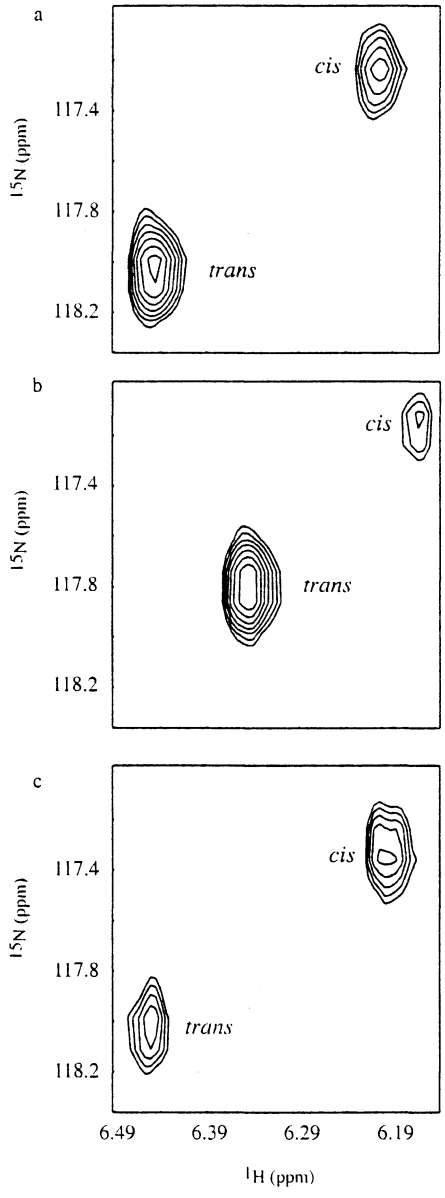


Figure 2-4. (a) Select region of the HSQC spectrum of 15N-labeled Itk SH2 showing the conformation-doubled resonances of Lys 258. The peaks corresponding to the cis and trans imide bond-containing conformations are labeled. (b) Corresponding region for 15N-labeled Itk SH2 domain following addition of a 9-fold excess of phosphopeptide. (c) Corresponding region for 15N-labeled Itk SH2 domain following addition of a 7-fold excess of Itk SH3 domain.

The proline-dependent conformational switch of Itk SH2 represents a novel form of control over protein-protein interactions. It is difficult at this point to say how widespread the existence of proline switches may be[1]. However, this type of reversible, noncovalent structural modification could provide an organism with significant advantages in the creation of dynamic signaling networks of the type that Itk appears to be involved in.
The full analysis of these networks requires a quantitative understanding of the affinity of protein-ligand interactions and the extent to which those affinities are regulated by conformational changes. This information cannot be obtained from the qualitative analysis that has been accomplished thus far. The development of methods to analyze the effect of conformational exchange on ligand binding in a quantitative manner could therefore prove a valuable tool in understanding molecular switches.

Experimental Methods

Data collection

Protein expression and purification

cDNA for full length Itk was supplied by the laboratory of Dr. Leslie Berg. Constructs coding for the SH2 and SH2 domains were amplified by PCR and ligated into pGEX-2T plasmids, where they would be expressed as glutathione-S-transferase (GST) fusion proteins. DNA sequencing performed by the Iowa State University DNA Sequencing and Synthesis Facility verified the accuracy of all constructed plasmids used in this research.

Protein domains were expressed as GST-fusions in the *E. coli* strain BL21(DE3). LB media was used to grow unlabeled proteins, whereas modified M9 minimal media containing ^{15}N-enriched ammonium chloride was used to generate ^{15}N-labeled proteins. When the *E. coli* growth had reached mid-log phase, as verified by optical density measurements at 600

nm, the temperature was reduced from 37°C to 15°C and protein expression induced by addition of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Expression was then allowed to proceed for 24 hours. Cells were then collected by centrifugation and resuspended in potassium phosphate buffer (50 mM potassium phosphate, 75 mM sodium chloride, 2 mM dithiothreitol, dissolved in electrolyte-pure water plus 0.002% sodium azide, adjusted to pH 7.4 using 3M sodium hydroxide). Lysozyme was then added and the cells stored at -80°C for at least 12 hours.

Cells were lysed upon thawing to room temperature. Protease inhibitor (PMSF) and DNase were added, and the cell lysate separated by centrifugation. The resulting supernatant was purified using a glutathione-agarose column. After extensive washing, the fusion protein was then eluted using 10 mM glutathione in potassium phosphate buffer. Proteins were concentrated and loaded onto a gel filtration column (Sephacryl S-100 HR). The protein-containing fractions were then pooled and cleaved at room temperature with thrombin for 3 hours (Itk SH2) or 12-16 hours (Itk SH3). Following cleavage, the proteins were passed over the glutathione column to remove GST, and once again concentrated and loaded onto the gel filtration column.

Protein-containing fractions were then checked for purity by SDS-PAGE and concentrated to final NMR levels. The concentration of the protein stocks was obtained by measuring UV absorption at 280 nm. All samples were stored at 5°C.

Peptide synthesis and purification

The phosphotyrosine-containing peptide ADpYEPPPSNDE was synthesized by the Iowa State University Protein Facility using standard solid-phase peptide synthetic techniques. The peptide was then purified using reversed-phase high performance liquid chromatography (HPLC) and lyophilized. The peptide was then resuspended in electrolyte-

pure water and its concentration measured using amino acid analysis, also carried out by the Iowa State University Protein Facility.

NMR spectroscopy

NMR spectra were acquired using a Bruker DRX500 spectrometer operating at a ^{1}H frequency of 499.867 MHz. A 5 mm triple resonance ($^{1}H/^{13}C/^{15}N$) probe with XYZ field gradients was used for all experiments. All NMR spectra recorded at 25°C. Gradient-enhanced HSQC experiments[8] were used for all spectra. Sequential assignments of both SH2 and SH3 domains were previously completed[6,9]. Data were acquired and processed using XWinNMR (Version 2.1, Bruker Instruments).

Data analysis

The NMRView[10] software package was used to view the NMR spectra in order to observe chemical shift changes and peak volumes. Peak volumes were measured using the elliptical volumes algorithm of NMRView. Determination of peak centers as well as volume integration boundaries were performed manually. Matlab (version 5.3.1, The Mathworks Inc.) was used for all equation fitting. Matlab scripts detailing the equation fitting procedures and algorithms used are available in Appendix B.

Results and Discussion

Measurement of equilibrium constants using NMR

The simplest model for SH2 binding to a ligand is given by:

$$P + L \underset{}{\overset{K}{\rightleftharpoons}} PL \qquad (1)$$

where P is the protein (Itk SH2), L the ligand (either phosphopeptide or Itk SH3), PL the protein/ligand complex, and K the association constant. The existence of well-resolved HSQC peaks that are affected by ligand binding allows for the measurement of chemical shift changes in the NMR tube upon addition of ligand. The normalized chemical shift of a peak, $|\delta_{obs}-\delta_{free}|/|\delta_{bound}-\delta_{free}|$, or $\Delta\delta/\Delta\delta_{max}$, is a direct measurement of f_b, the fraction of SH2 in the ligand-bound state. This can then be used to extract the association constant K using the following equation[11]:

$$\Delta\delta / \Delta\delta_{max} = f_b = \frac{P_0 + L_0 + K_o^{-1} - \sqrt{(P_0 + L_0 + K_o^{-1})^2 - 4P_0L_0}}{2P_0} \qquad (2)$$

where P_0 and L_0 represent the total concentrations of protein (Itk SH2) and ligand (Itk SH3 or phosphopeptide). K and $\Delta\delta_{max}$ are the only unknown parameters in the above equation, and both are optimized in the fitting of data to (2).

In addition to the requirements for reporter peaks to be well-resolved and to display measurable chemical shift differences between free and bound states, it is also necessary for the peak to be unaffected by cis/trans isomerization. The reason for this is that the cis and trans conformers of the peptide undergo interconversion, and are allowed, therefore, to vary in concentration during the course of a titration.

Based on these requirements, five SH2 HSQC peaks (N241, Y292, V311, F312, and L318) were used to determine the binding affinity of phosphopeptide to Itk SH2 (Figure 2-5 a). In the case of Itk SH3 binding, no Itk SH2 peaks were found that fit all the necessary requirements. The observed association constant between SH2 and SH3 was therefore measured using the reverse labeling scheme (unlabeled SH2 domain titrated into ^{15}N-labeled SH3 domain). The concentration dependence of the normalized chemical shifts for five SH3 HSQC peaks (Y180, W208, W209, A221, and S223) are shown in Figure 2-5 b. The association constant for each binding event is reported in Table 2.

Throughout a titration, the total concentrations of protein and ligand are known; therefore, once the association constant has been determined, f_b can be calculated at each point using (2). Meanwhile, the integrated volumes of the cis and trans peaks for conformational exchange-doubled residues give a direct measurement of f_c, the fraction of the SH2 domain in the cis conformation. The relationship between f_c and f_b is linear (derivation in later section). Therefore, if the Itk SH3 domain binds exclusively to the cis Itk SH2 conformer, then at $f_b = 1$, f_c will also equal 1. Similarly, for the phosphopeptide titration, at $f_b = 1$ a value of $f_c = 0$ is be expected if only the trans conformer is able to bind the peptide. This, however, is not the case for either the phosphopeptide or the SH3 titrations (Figure 2-6, a and b, respectively). This observation indicates that both SH2 conformers have measurable affinities for each of the SH2 ligands, and shows that a new SH2 ligand-binding model must be developed which takes this into account.

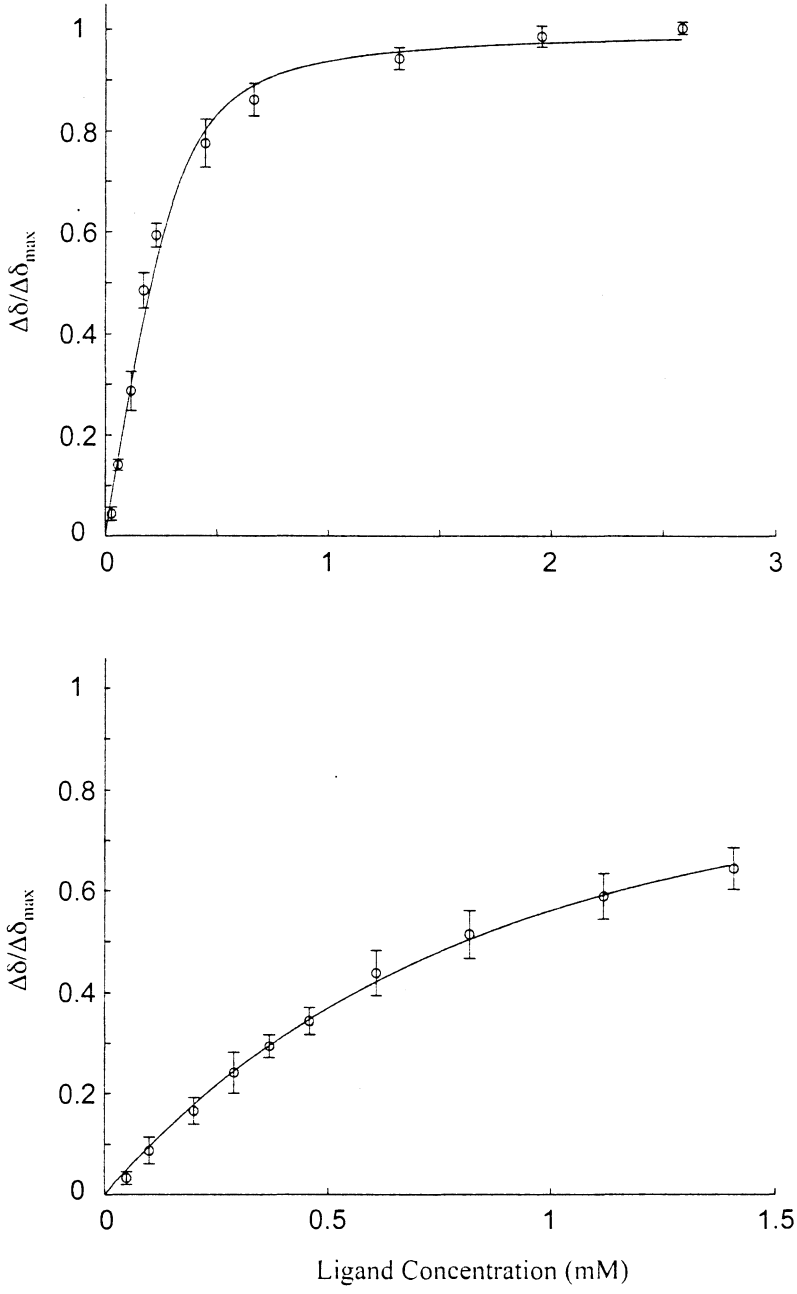
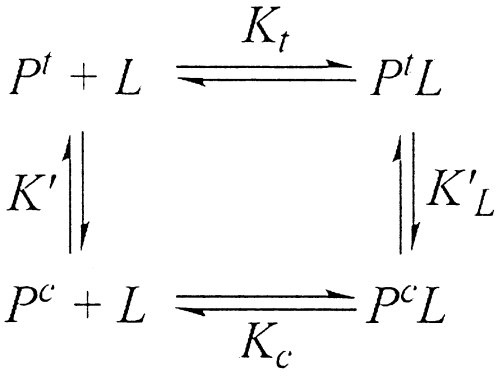


Figure 2-5. Concentration dependence of normalized chemical shifts for (a) titration of phosphopeptide into ^{15}N-labeled Itk SH2 domain (0.32 mM) and (b) titration of unlabeled Itk SH2 domain into ^{15}N-labeled Itk SH3 domain (0.32 mM). In both cases, the curves depict the best fit of (2) to the NMR data. Error bars represent two standard deviations about the mean.

The cis/trans ligand-binding model

Based on these findings, the following equilibrium model was proposed for each SH2 ligand:



where the *c* and *t* subscripts refer to the cis and trans imide-bond containing conformers, respectively. It is not necessary, however, to include all four of the above equilibrium constants in a mathematical description of the system. One of the constants may be rewritten in terms of the other three (see Derivations of the governing equations). While to some extent, the question of which equilibrium constants to work with is arbitrary, I have decided to mathematically model the system using K', K_c, and K_t, leaving out K'_L. This decision was made for the following three reasons: (a) It is unknown whether the cis/trans isomerization of ligand-bound SH2 occurs directly in solution, (b) the focus of this work is the effect of isomerization upon binding affinity, rather than the effect of ligand binding upon isomerization, and (c) unlike K', K'_L cannot be directly measured experimentally. Regardless of how the model is written, the results will be the same. The calculated values of K'_L for each ligand are reported in the caption of Table 2-2.

It should be noted that the model presented above would be affected by the oligomerizaion of any of the species involved. The assumption that this does not occur is well justified due to the fact that NMR studies of Itk SH2 (K. Brazin & A. Andreotti,

unpublished results), Itk SH3[12], and phosphopeptide (M. Sundd & A. Andreotti, unpublished results) reveal no indication of oligomerization. For Itk SH2 in particular, a concentration dependence study was run on over the range 0.01 mM to 1mM, and no changes in the HSQC spectrum were observed. These experiments also establish that the cis/trans ratio of the Itk SH2 domain is not subject to concentration dependent effects.

Overview of the fitting process

The application of conventional binding assays to the study of SH2 behavior would only reveal an averaging of the behavior of the two conformations, thereby masking the ability to measure each conformer separately. The above model highlights the fact that the simple association constant represented in (1) and (2) is an inaccurate representation of the association of SH2 and ligand. For this reason, the association constant given by (2) is renamed from this point on K_o. The subscript refers to the fact that this is an "observed" association constant which does not take into account the two conformations of the SH2 domain, but rather blends the affinities of each conformer into an overall, observed, association constant.

By revealing the inadequacy of the observed association constant to describe the chemical equilibria, the cis/trans ligand-binding model also makes clear the need to measure independent ligand-binding affinities, K_c and K_t, for each conformer. Accomplishing this requires deriving equations from the model presented above which can be fit to accurately measurable NMR data in a manner that is both computationally feasible and yields consistent results, regardless of the residue used to quantify the equilibrium constant. In fitting equations to data, the reliability of the fit is greatly improved by limiting to as low a number as possible the number of unknown parameters that must be fit to the data. The method developed here takes the following approach for each ligand. First, K_o is fit to chemical shift data using the well-established method described by Equation 2. Second, conformer-ligand

affinities are fit to peak volume integration data using a single fit with a single unknown parameter. Both fits are accomplished using unconstrained nonlinear optimization algorithms.

As described earlier, the relationship between f_b and f_c is linear. This linear relationship depends on the relative ligand-binding affinities of the cis and trans conformations, K_c and K_t:

$$f_c = \frac{K'(K_c - K_t)}{(1+K')(K_t + K_c K')} f_b + \frac{K'}{(1+K')} \tag{3}$$

To increase the accuracy of the fit, K_o can be used to determine the relationship between K_c and K_t:

$$K_o = \frac{K_t + K_c K'}{1+K'} \tag{4}$$

so that (3) may be rewritten as a function involving only one unknown parameter and fit to the NMR volume integration data (choice of K_t or K_c is arbitrary – after the fit, (4) is used to determine the other association constant in a manner independent of which parameter is actually used in the fit and which one is calculated afterwards). The derivations of (3) and (4) are presented below.

Derivation of the governing equations

The cis/trans model as presented here is defined by these five equations:

$$K_t = \frac{[P^t L]}{[P^t][L]} \tag{5}$$

$$K_c = \frac{[P^c L]}{[P^c][L]} \tag{6}$$

$$K' = \frac{[P^c]}{[P^t]} \tag{7}$$

$$P_0 = [P^t] + [P^c] + [P^t L] + [P^c L] \tag{8}$$

$$L_0 = [L] + [P^t L] + [P^c L] \tag{9}$$

In the derivations that follow, the identities below, easily derived from Equations 5-7, prove useful:

$$[P^c] = K'[P^t] \tag{10}$$

$$[P^c L] = \frac{K_c K'}{K_t}[P^t L] \tag{11}$$

It is worth noting that (11) is equivalent to writing the model in a way that includes the K'_L term. It is apparent that K'_L may be rewritten as $K_c K'/K_t$ without a loss of generality.

Derivation of (3)

The quantities f_b and f_c are defined in this system as:

$$f_c = \frac{[P^c] + [P^c L]}{P_0} \tag{12}$$

$$f_b = \frac{[P^c L] + [P^t L]}{P_0} \tag{13}$$

Equation 3 consists of transforming Equation 12 into an equation linear in f_b.

To begin, both the numerator and denominator of Equation 12 can be multiplied by $(1 + K')(K_t + K_cK')$ to yield

$$f_c = \frac{K_t[P^c] + K_tK'[P^c] + K_cK'[P^c] + K_cK'^2[P^c] + K_t[P^cL] + K_tK'[P^cL] + K_cK'[P^cL] + K_cK'^2[P^cL]}{P_0(1+K')(K_t + K_cK')}$$

We can then factor the terms in the numerator into three groups by using (10) and (11):

$$f_c = \frac{K_tK'[P^t] + K_tK'[P^c] + K_cK'^2[P^t] + K_cK'^2[P^c] + K_cK'[P^tL] + K_cK'^2[P^tL] + K_cK'[P^cL] + K_cK'^2[P^cL]}{P_0(1+K')(K_t + K_cK')}$$

$$f_c = \frac{K_cK'([P^cL] + [P^tL]) + K_tK'([P^c] + [P^t]) + K_cK'^2 P_0}{P_0(1+K')(K_t + K_cK')} \tag{14}$$

where in Equation 14, the four variables that were associated with the constant $K_cK'^2$ are rewritten as P_0 by using (8). We can then use (8) to rewrite the K_tK' term and refactor variables:

$$f_c = \frac{K_cK'([P^cL] + [P^tL]) + K_tK'(P_0 - [P^cL] - [P^tL]) + K_cK'^2 P_0}{P_0(1+K')(K_t + K_cK')}$$

$$f_c = \frac{K'(K_c - K_t)([P^cL] + [P^tL]) + K'(K_t + K_cK')P_0}{P_0(1+K')(K_t + K_cK')}$$

$$f_c = \frac{K'(K_c - K_t)([P^cL] + [P^tL])}{P_0(1+K')(K_t + K_cK')} + \frac{K'(K_t + K_cK')P_0}{P_0(1+K')(K_t + K_cK')}$$

$$f_c = \frac{K'(K_c - K_t)}{(1+K')(K_t + K_cK')} f_b + \frac{K'}{(1+K')} \tag{15}$$

Derivation of (4)

In the system given by (5)-(9), K_o must be rewritten in order to account for *cis/trans* isomerization:

$$K_o = \frac{[P^t L] + [P^c L]}{([P^t] + [P^c])[L]} \tag{16}$$

The right side of (16) can be rewritten using (10) and (11) to yield:

$$K_o = \frac{[P^t L] + \frac{K_c K'}{K_t}[P^t L]}{([P^t] + K'[P^t])[L]}$$

$$K_o = \frac{[P^t L](1 + \frac{K_c K'}{K_t})}{[P^t][L](1 + K')} \tag{17}$$

Then, using (5), the right side of (17) becomes

$$K_o = K_t \frac{1 + \frac{K_c K'}{K_t}}{1 + K'}$$

$$K_o = \frac{K_t + K_c K'}{1 + K'} \tag{18}$$

Application of the cis/trans ligand-binding model to Itk SH2

If quantitative claims are to be made regarding the relative concentration of cis and trans conformers in solution on the basis of residue signal intensity, the effect of differential relaxation rates on intensity must be analyzed. As discussed in Chapter 1, in addition to the

requirement of giving rise to a pair of isolated peaks, residues must also be chosen that do not have significantly different relaxation parameters between their cis and trans conformations.

To ensure that differential relaxation effects would not affect the analysis, ^1H longitudinal relaxation times were measured for Itk SH2 using inversion recovery experiments. Table 2-1 reports the measured relaxation parameters for the residues used in the subsequent analysis. The table also illustrates that differences in the relaxation times between cis and trans states of the residues used in the analysis will not cause significant error in the measured cis/trans ratio in these experiments. Even the largest errors are well within experimental error, and considerably smaller than the relative differences in affinity between cis and trans forms for the SH2 ligands.

In order to fit (3) to the NMR data for Itk SH2 and its ligands, five residues were selected which fit the requisite criteria of peak isolation and relaxation uniformity. The volumes of the conformation-doubled HSQC peaks of these residues were then integrated over the course of a ligand titration. The trans to cis interconversion constant K' of free SH2 is given directly by the integrated volume ratio of the cis and trans cross peaks (Table 2-2). K_c and K_t were then obtained by fitting (3) to the peak integral data of five cis/trans pairs (Figure 2-6). The solid lines in Figure 2-6 represent the fit of (3) to the NMR data. The values of K_c and K_t, as determined by the fit, are given in Table 2-2, along with the observed association constants for both ligand titrations.

Table 2-1. ^{1}H T_1 relaxation parameters for Itk SH2 residues used in the analysis of cis/trans population ratio. The magnetization recovered between scans (I/Io) is dependent on the ^{1}H T_1 relaxation of the backbone amide as well as the relaxation delay used in the experiment. For all experiments in this chapter, a relaxation delay of 2 seconds was used. The largest error due to differential relaxation is expected for Lys 290, where I/Io differs by 5.9% between the cis and trans forms. All measurements were conducted on 1mM SH2 at 298K.

	^{1}H T_1 (ms)		$I/I_0=\exp(-T_1/RD)$	
Residue	Cis	trans	cis	trans
256	896±40	866±40	0.638	0.649
257	1155±40	1160±40	0.561	0.560
258	936±40	924±70	0.626	0.630
260	1024±70	981±30	0.599	0.612
278	1046±70	1052±70	0.593	0.591
288	806±40	828±30	0.668	0.661
290	998±30	876±40	0.607	0.645

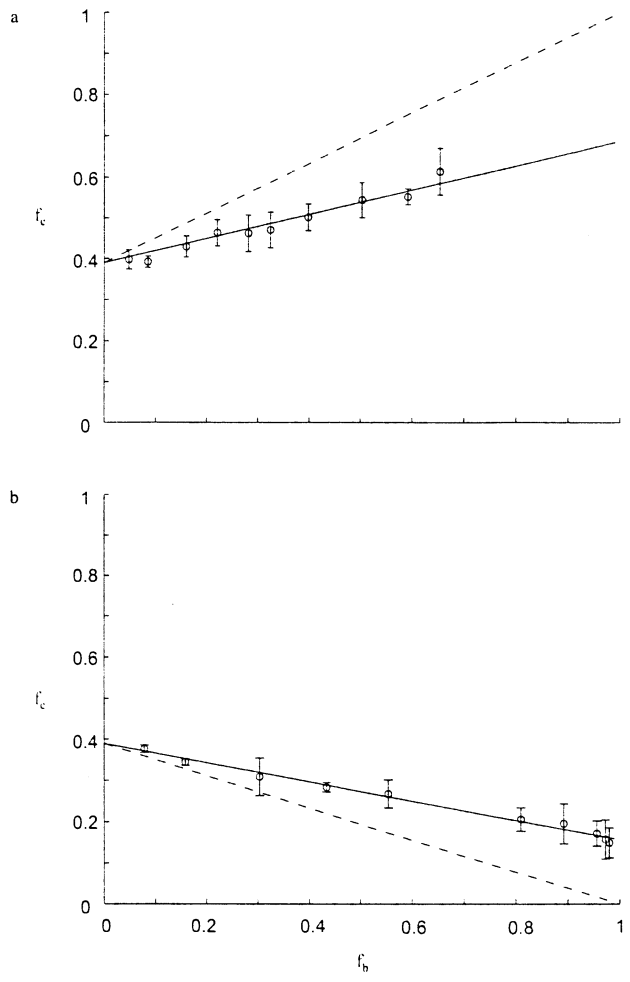


Figure 2-5. Plot of the fraction of Itk SH2 in the cis conformation (f_c) versus the total fraction of SH2 bound to ligand (f_b) for (a) titration of phosphopeptide into a sample containing ^{15}N-labeled Itk SH2 domain (0.32 mM), and (b) titration of the Itk SH3 domain into a sample containing ^{15}N-labeled Itk SH2 domain (0.32 mM). Residues T256, K258, G257, F278, and G260 were used in (a), while residues C288, K258, G257, K290, and G260 were used in (b). Error bars represent two standard deviations about the mean. The solid line represents the best fit of (3) to the data. The dotted line represents the expected fraction of cis if phosphopeptide exclusively bound the trans conformer ($K_c = 0$) in (a) or Itk SH3 exclusively bound the cis conformer ($K_t = 0$) in (b).

Table 2-2. Observed, cis, and trans association constants for Itk SH2 binding to each of its ligands. The error represents the standard deviation between the five residues used to measure the association constant. The actual uncertainty in the measurement is larger, including contributions from any systematic sources of error. In addition, the uncertainty in the measurement of K_t and K_c is affected by the uncertainty of the Ko measurement, which is used to calculate K_t and K_c. The value for K' as measured in this method is 0.63 ± 0.07. The value for K'_L (phosphopeptide) is 0.19 ± 0.02, while the value for K'_L (SH3) is 2.2 ± 0.2.

	Itk SH2 binding phosphopeptide[a] (mM^{-1})	Itk SH2 binding Itk SH3 (mM^{-1})
K_o	21 ± 1	1.5 ± 0.1
K_t	29 ± 0.4	0.7 ± 0.1
K_c	8.4 ± 0.7	2.6 ± 0.1

[a] Constants reported here for the SH2-phosphotyrosine binding differ from published values[13] due to different methods used to quantitate the peptide concentration. The published values used gravimetric methods to determine the peptide concentration, whereas the values reported here relied on amino acid analysis, as described in the materials and methods sections. Due to the inaccuracy of measuring small volumes, amino acid analysis was judged to give a more accurate measure of phosphopeptide concentration. It is worth noting, however, that both these values and the published values yield approximately the same ratio between cis and trans affinities (3.5 and 4, respectively).

Conclusions

The method developed in this chapter shows that a quantitative analysis of the NMR data reveals the extent to which ligand-binding affinities are modulated by the isomerization between Itk conformations. While it was revealed that conformational exchange does not exclude binding of each ligand to its unpreferred conformer, each conformer exhibited a 3.5-fold higher affinity for its respective ligand than the alternate conformation. Thus, for the Itk SH2 domain, cis/trans isomerization of a single prolyl imide bond affords this small domain the ability to control the relative binding affinities for distinct ligands during cell signaling.

Despite great progress in the field over the last half-century, a great deal remains unknown regarding the structural mechanisms by which proteins recognize and modulate binding to their cellular targets. These precise structural questions are unaddressed by most biochemical experiments. Nevertheless, the importance of protein-protein interaction domains in modulating protein activity has emerged, particularly in cell signaling pathways. The mechanisms by which these domains alter their ligand-binding affinities may act to regulate the enzyme directly or affect the protein through binding of cellular targets.

Proline isomerization, which is particularly difficult to detect by methods other than NMR, is a possibly widespread mechanism by which the modulation of ligand affinity may occur. The quantitative understanding of the effect of proline isomerization on ligand specificity may therefore prove to be an important matter in understanding the mechanisms by which the diversity and regulation of protein-protein interactions is achieved.

References

1. Andreotti A. H. Native state proline isomerization: an intrinsic molecular switch. *Biochemistry* **42**, 9515-9524 (2003).
2. Wright P. E. & Dyson H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paridigm. *J. Mol. Biol.* **293**, 321-331 (1999).
3. Mallis R. J., Brazin K. N., Fulton D. B., & Andreotti A. H. Structural characterization of a proline-driven conformational switch within the Itk SH2 domain. *Nat. Struct. Biol.* **9**, 900-905 (2002).
4. Wüthrich K., Billeter M., & Braun W. Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J. Mol. Biol.* **180**, 715-740 (1984).
5. Brazin K. N., Mallis R. J., Fulton D. B., Andreotti A. H. Regulation of the tyrosine kinase Itk by the peptidyl-prolyl isomerase cyclophilin A. *Proc. Natl. Acad. Sci.* **99**, 1899-1904 (2002).
6. Brazin K. N., Fulton D. B., & Andreotti A. H. A specific intermolecular association between the regulatory domains of a Tec family kinase. *J. Mol. Biol.* **302**, 607-623 (2000).
7. Koradi R., Billeter M., & Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structure. *J. Mol. Graph.* **14**, 29-32 (1996).
8. Kay L. E., Keifer P., & Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663-10665 (1992).
9. Andreotti A. H., Bunnell S. C., Feng S., Berg L. J., & Schreiber S. L. Regulatory intramolecular association in a tyrosine kinase of the Tec family. *Nature* **385**, 93-97 (1997).

10. Johnson B. A. & Blevins R. A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603-614 (1994).

11. Lian L. Y. & Roberts G. C. Effects of chemical exchange on NMR spectra. In *NMR of Macromolecules: A Practical Approach* (Roberts GC, ed.), pp. 153-182. Oxford University Press, Inc., New York.

12. Laederach A., Cradic K. W., Fulton D. B., & Andreotti A. H. Determinants of intra versus intermolecular self-association within the regulatory domains of Rlk and Itk. *J. Mol. Biol.* **329**, 1011-1020 (2003).

13. Breheny P. J., Laederach A., Fulton D. B., & Andreotti A. H. Ligand specificity modulated by prolyl imide bond cis/trans isomerization in the Itk SH2 domain: a quantitative NMR study. *J. Am. Chem. Soc.* **125**, 15706-15707 (2003).

CHAPTER 3: FUTURE DIRECTIONS AND CONCLUSION

Future Directions

The method presented in Chapter 2 provides what is essentially a complete description of the extent to which the conformational heterogeneity of the Itk SH2 domain affects binding affinity to Itk SH3 and phosphopeptide. The logical extension of the ideas expressed in that chapter would be the application of the method to the study of more complex and biologically relevant systems. The obvious next step is examination of the effect of SH2 isomerization upon the dimerization of the SH3-SH2 fragment. Initial work, however, shows that the SH3SH2 fragment must be understood to a greater extent before such a detailed analysis can take place. Recent efforts have thus concentrated on characterizing the dimerization of this SH3SH2 dual domain in terms of the affinity of the association and the determination of its topology, before moving on to answer the question of the role of the proline switch in dimerization.

Measuring Affinity

The difficulty in applying the methods presented in chapter 1 to SH3SH2 lie in the fact that one cannot measure the affinity of the dimerization reaction in the same way as the single-domain interactions. Self-association affinity is typically measured by varying the concentration of the self-associating species. Because NMR requires high concentrations in order to achieve a good signal-to-noise ratio, the range of concentrations for which one can acquire clear spectra is quite limited. Therefore, attention has turned to measuring the association constant via other techniques. An approximate $K_{a(dimer)}$ (association constant for dimerization) of 40 mM^{-1} has been reported earlier using analytical ultracentrifugation[1]. Currently, the most attractive technique for deriving a reliable and quantitative measurement

of $K_{a(dimer)}$ for SH3SH2 seems to be the use of dynamic light scattering to monitor the dependence of the observed molecular weight on concentration. Initial efforts to conduct this study resulted in very noisy data, a result possibly due to interference from dust in the sample. This study is scheduled to be repeated taking additional measures to filter dust from the sample.

An alternative approach that still allows the use of NMR is to fix the SH3SH2 concentration and obtain the dimerization affinity by titrating in a peptide ligand that will bind one of the domains in such a way that the dimer is dissociated. By calibrating the affinity of the peptide for the single domain, the ability of the peptide to dissociate the dimer can be used to gauge the inherent affinity of the monomers for each other.

There is, however, a difficulty associated with this approach as well. In fitting association curves to the SH2 domain using $\Delta\delta/\Delta\delta_{max}$, the chemical shift of free SH2 domain was known, and therefore K_a and δ_{bound} were the only two unknowns. In the case of dimerization, neither δ_{bound} nor δ_{free} is known, and in initial efforts to extract K_D from such a competitive-peptide titration, this factor has proven significant in diminishing the reliability of the fit.

Determining topology

Another important question regarding the self-association of the SH3SH2 dual domain is the topology of the dimerized structure. Without knowing the structure of the entire SH3SH2 fragment, it is difficult to predict whether a symmetric, fully-occupied dimer is likely to form or if such a dimer is sterically unfeasible and the mode of self-association is an asymmetric, half-occupied dimer. A schematic of the two modes is presented in Figure 3-1. The nature of the dimer topology will impact any mathematical model which attempts to assess of the impact of the proline switch in the SH2 domain on dimerization.

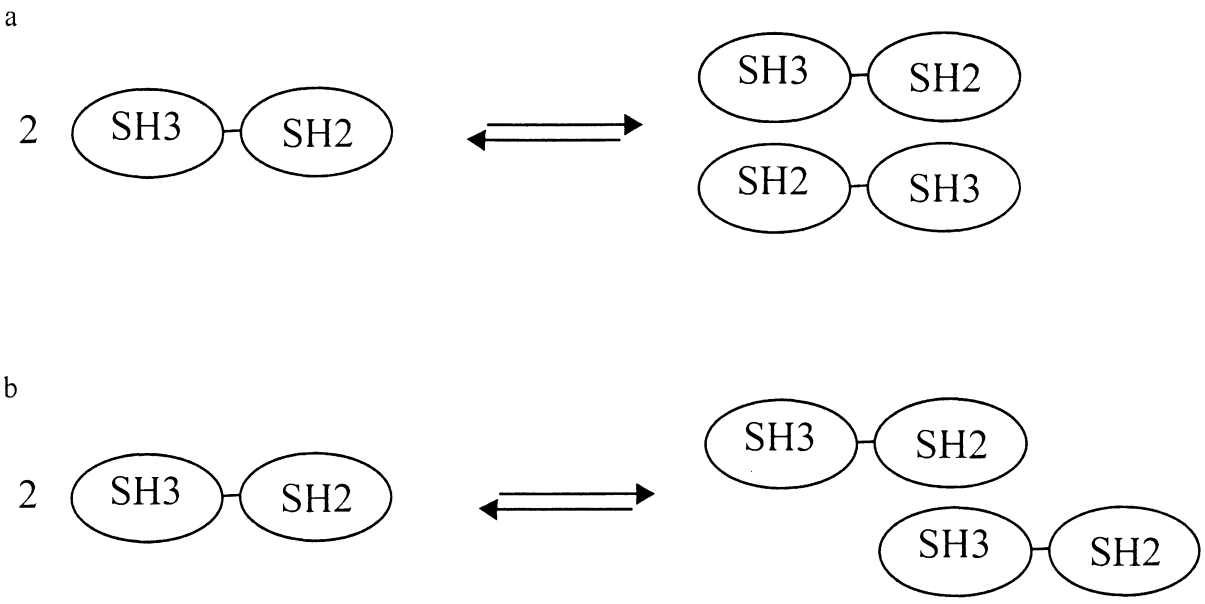


Figure 3-1. Possible topological states of the SH3SH2 dimer. (a) A symmetric, fully-occupied dimer. (b) An asymmetric, half-occupied dimer.

This question is being probed using compensatory mutants. Itk SH3SH2 (W208K) is a mutant developed by the laboratory of Leslie Berg that abrogates all intermolecular interactions with the SH3 domain binding groove. NMR studies reveal that the attached SH2 domain is unaffected and that the construct behaves in an entirely monomeric fashion. Our lab is attempting to develop a similar mutation in the SH2 domain that eliminates binding of SH3. Mutants developed thus far precipitate at high concentrations, and this question has thus proven difficult to answer using NMR. In addition to developing and screening new mutants, efforts are also underway to measure the ability of the compensatory mutants to bind each other either using other techniques.

While the behavior of the Itk SH3 and SH2 domains is well understood in experiments involving the isolated domains, the larger SH3SH2 construct remains insufficiently characterized. A better understanding of this fragment and its self-associating

behavior would give much-needed insight into the nature of Itk dimerization, the functional significance of which in the full-length protein remains very unclear.

Conclusion

As a conclusion to this thesis, I feel it is important to address the question of what the significance of this work is beyond what was previously understood. To answer that question, I would present the following main points regarding the significance of the completed work presented here.

- Cis/trans isomerization does not confer full-ligand binding specificity to the Itk SH2 domain. Figure 2-6 convincingly shows that while cis/trans isomerization certainly imparts specificity to its associated conformers, Itk SH2 in the cis conformation is still capable of binding phosphopeptide, and Itk SH2 in the trans conformation is capable of binding Itk SH3. Therefore, either more interactions are required to confer additional specificity, or the unpreferred interactions are physiologically relevant. Either way, mechanistic models of the activation of Itk must take this finding into account.
- In addition to the Itk specific findings, these results are of general interest in gaining a greater quantitative understanding of the magnitude of effects that proline switches may be capable of.
- The development of this methodology for independently measuring the independent ligand-binding association constants for multiple distinct protein conformations in a solution could be of general use in studying other such systems. While in theory this method is applicable to any conformationally heterogeneous protein, it is unclear whether conformational exchange events exist other than proline isomerization that would occur slowly enough to give rise to slow exchange NMR peaks, and yet fast

enough that physical separation of the two conformations is not feasible. Essentially then, the significance of this method is tied to the significance of proline switches. If in fact, proline switches are a widespread form of molecular control, then the quantitative study and understanding of them will prove important, and this type of analysis may prove a useful tool in probing them.

References

1. Brazin K. N., Fulton D. B., & Andreotti A. H. A specific intermolecular association between the regulatory domains of a Tec family kinase. *J. Mol. Biol.* **302**, 607-623 (2000).
2. Mallis R. J., Brazin K. N., Fulton D. B., & Andreotti A. H. Structural characterization of a proline-driven conformational switch within the Itk SH2 domain. *Nat. Struct. Biol.* **9**, 900-905 (2002).
3. Young M. A., Gonfloni S., Superti-Furga G., Roux B., & Kuriyan J. Dynamic coupling between the SH2 and SH3 domain of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115-126 (2001).

ACKNOWLEDGEMENTS

Two years ago, I didn't even know what proteins were. So obviously, I've had a lot of help along my way, and I would like to thank many people, without whose help I would be a very different person today. First and foremost, I would like to thank Amy Andreotti and Alain Laederach for their patience and wisdom in teaching me how to be a scientist. They were the best mentors I could ever have hoped for. I would also like to thank Bruce Fulton and Kate Pletneva for teaching me about NMR and protein purification, respectively; Melissa Mayo for being a great friend and for numerous enlightening conversations; Robbie McGinty and Lie Min for being tremendous role models of productivity, generosity, and integrity; Monica Sundd and Andrew Severin for their contributions and collaboration with me in my research; and Iowa State University, for providing me with a wonderful academic environment.

APPENDIX: MATLAB SCRIPTS

Four MATLAB scripts are included in this appendix:

- fitproteinligandcurve.m (fits an association constant to a protein-ligand titration).
- fitproteinliganderror.m (called by fitproteinligandcurve; generates a sum of squares error for comparing observed data to theoretical prediction).
- fitmlciscurve.m (fits association constants for cis and trans conformers to the cis-trans ligand-binding model).
- fitmlciserror.m (called by fitmlciscurve; generates a sum of squares error for comparing observed data to theoretical prediction).

```
%This function fits an association constant to a protein-ligand titration
%and plots a binding curve depicting the association.

%Conc = Array containing the concentration of protein and ligand as they
%vary throughout the titration.
%DeltaArr = Array containing the imported chemical shifts.
%Start;End = Peaks to which the data should be fit (i.e. columns in
%DeltaArr that should be used by the program).
%KaEst = Rough estimate of the Ka of the interaction, in mM^(-1)

%KaArr = Array containing association constants as determined by each peak
%used in the fit.
%KaAve = Average of the association constants contained in KaArr.
%KaErr = Standard deviation of the association constants containted in
%KaArr.
%DataPoints = Array containing the Y axis values for the data points
%plotted in the binding curve.


function [KaArr, KaAve, KaErr, DataPoints, DeltaBound] =
fitproteinligandcurve(Conc, DeltaArr, Start, End)


%PART I: PROCESSING THE RECEIVED VARIABLES
ProteinConc = Conc(:,2);
LigandConc = Conc(:,1);

Points = size(ProteinConc);
Points = Points(1);
%Points = Number of points in the titration

Options = optimset('Display','iter');
%Required by fminsearch to display optimization output
```

```
Value = 1;
%Needed to ensure KaArr, KbArr start at 1 even if Start is not equal to 1

%Difference = DeltaArr(Points,:) - DeltaArr(1,:);
%DeltaBound = DeltaArr(1,:) + Difference;


%PART II: OPTIMIZING THE FIT
for i = Start:End
    StartParams(1) = 1;
    StartParams(2) = DeltaArr(Points, i);
    %StartParams = Initial values for Ka, DeltaBound that will be
optimized
    %by fminsearch

    Params = fminsearch('fitproteinliganderror', StartParams, Options,
Conc(2:Points, :), DeltaArr(2:Points, i), DeltaArr(1,i));
    KaArr(Value) = Params(1);
    DeltaBound(Value) = Params(2);
    %Params, KaArr, DeltaBound = Optimized values yielded by fminsearch

    for j = 2:Points
        DataPoints(j, Value) = (abs(DeltaArr(j,i) -
DeltaArr(1,i)))/(abs(DeltaBound(Value) - DeltaArr(1,i)));
    end

    Value = Value + 1;
end

KaAve = mean(KaArr);
KaErr = std(KaArr);


%PART III: PLOTTING THE DATA POINTS
hold off;
axis([0 4.5 0 1.05]);
hold on;
%Creates a figure window and determines the x and y axes of the plot

for i = 1:(Points - 1)
    DataPointsMean(i) = mean(DataPoints(i + 1,:));
    Error(i) = 2*std(DataPoints(i + 1,:));
end

%Optional: This section can be decommented and the preceding section
%commented in order to toggle between displaying the binding curve as a
%scatter plot or with error bars.
%for i = 1:(Value - 1)
%    plot(Conc(2:Points,1), FittedCurveArr(2:Points,i),'ko');
%    hold on;
%end

errorbar(Conc(2:Points,1), DataPointsMean, Error, 'ko');
hold on;
%Plots the data points as black circles
```

```
%PART IV: PLOTTING THE BINDING CURVE
ProteinConcMin = min(Conc(:,2));
ProteinConcMax = max(Conc(:,2));

LigandConcMax = max(Conc(:, 1));
LigandConc = 0.0001:0.01:LigandConcMax;

Points = size(LigandConc);
Points = Points(2);
%Points = Number of points for which the value of the value of the binding
%curve is calculated

for i = 1:Points
    ProteinConc(i) = ProteinConcMax - (ProteinConcMax -
ProteinConcMin)*i/Points;
end
%Protein concentration is assumed to vary linearly with ligand
%concentration over the course of the titration

for i = 1:Points
    a2 = KaAve;
    a1 = -KaAve*ProteinConc(i) - KaAve*LigandConc(i) - 1;
    a0 = KaAve*ProteinConc(i)*LigandConc(i);

    FittedConcPL(i) = (-a1 - sqrt(a1*a1 - 4*a2*a0))/(2*a2);
    FittedConcL(i) = LigandConc(i) - FittedConcPL(i);
    FittedConcP(i) = ProteinConc(i) - FittedConcPL(i);
    %Concentrations of protein-ligand complex, free ligand, and free
    %protein, respectively

    FittedCurve(i) = FittedConcPL(i)/ProteinConc(i);
end

plot(LigandConc, FittedCurve,'k');
```

```
%This function returns a value corresponding to the SOS error of fitting a
%simple protein-ligand binding model to the data given using the passed
%values for bound chemical shift and association constant.

%Params = Parameters being minimized by fminsearch (Kd and DeltaBound)
%Conc = Array containing the concentration of protein and ligand as they
%vary throughout the titration (unlike other programs, Conc here does not
%include the first point of the titration).
%DeltaArr = Array containing the imported chemical shifts (again, does not
%include here the first titration point).
%DeltaFree = Rough estimate of the Ka of the interaction, in mM^(-1)

%Error = Sum of the squares of the difference between the value of the
%fitted curve at a titration point and the experimentally observed value.

%2003 Patrick Breheny.

function Error = fitproteinliganderror(Params,Conc,DeltaArr,DeltaFree)

KaEst = Params(1);
DeltaBoundEst = Params(2);
%DeltaBoundEst = DeltaBound;
%KaEst = Estimated value of the association constant.
%DeltaBoundEst = Estimated value of the chemical shift for the protein
%bound to ligand.

ProteinConc = Conc(:,2);
LigandConc = Conc(:,1);

Points = size(ProteinConc);
Points = Points(1);
%Points = Number of titration points (not including the first, which has
no
%ligand added).

for i = 1:Points
    a2 = KaEst;
    a1 = -KaEst*ProteinConc(i) - KaEst*LigandConc(i) - 1;
    a0 = KaEst*ProteinConc(i)*LigandConc(i);

    FittedConcPL(i) = (-a1 - sqrt(a1*a1 - 4*a2*a0))/(2*a2);
    FittedConcL(i) = LigandConc(i) - FittedConcPL(i);
    FittedConcP(i) = ProteinConc(i) - FittedConcPL(i);
    %Concentrations of protein-ligand complex, free ligand, and free
    %protein, respectively

    FittedCurve(i) = DeltaFree*FittedConcP(i)/ProteinConc(i) +
DeltaBoundEst*FittedConcPL(i)/ProteinConc(i);
end

for i = 1:Points
   ErrorVector(i) = FittedCurve(i) - DeltaArr(i);
end

Error = sum(ErrorVector.^2);
%This function fits association constants for the cis and trans forms of a
%confomationally heterogenous protein to a FractionCis/FractionBound curve
```

```
%based on an observed Ka and compares it to a model in which only the cis
%conformation binds.

%Conc = Array containing the concentration of protein and ligand as they
%vary throughout the titration.
%CVol, TVol = Array containing the integrated volumes of cis and trans
%peaks, respectively, as they vary throughout the titration.
%Start;End = Peaks to which the data should be fit (i.e. columns in
%CVol, TVol that should be used by the program).
%Ko = Observed Ka of the interaction, ignoring isomerization.
%KPrime = Interconversion constant between the two isomers.

%KCisArr, KTransArr = Array containing association constants as determined
%by each peak used in the fit.
%KCisAve, KTransAve = Average of the association constants contained in
%KCiSArr, KTransArr.
%KCisErr, KTransErr = Standard deviation of the association constants
%contained in KCisArr, KTransArr.
%DataPoints = Array containing the Y-axis values for the data points
%plotted in the binding curve.


function [KCisArr, KTransArr, KCisAve, KTransAve, KCisErr, KTransErr,
DataPoints] = fitmlciscurve(Conc, CVol, TVol, Start, End, Ko, KPrime)


%PART I: Processing the received variables
ProteinConc = Conc(:,2);
LigandConc = Conc(:,1);

Points = size(ProteinConc);
Points = Points(1);
%Points = Number of points in the titration

for i = 2:Points
    FractionBound(i - 1) = (ProteinConc(i) + LigandConc(i) + (1/Ko) -
sqrt(((ProteinConc(i) + LigandConc(i) + (1/Ko))^2) -
4*ProteinConc(i)*LigandConc(i)))/(2*ProteinConc(i));
end
FractionCis = CVol(2:Points,:)./(TVol(2:Points,:) + CVol(2:Points,:));
%FractionBound = Fraction of total protein bound to ligand.  This is
%calculated at each point in the titration based on the observed Ka of the
%association.
%FractionCis = Fraction of total protein in cis conformation.

Options = optimset('Display','iter');
Value = 1;          %Needed to ensure KaArr, KbArr start at 1 even if
Start is not equal to 1


%PART II: Optimizing the fit
for i = Start:End
    StartParam = Ko;
    %StartParam = Initial value for KCis that will be optimized by
    %fminsearch
```

```
    Param = fminsearch('fitmlciserror', StartParam, Options,
FractionBound, FractionCis(:,i), KPrime, Ko);
    KCisArr(Value) = Param;
    KTransArr(Value) = Ko*(1 + KPrime) - KCisArr(Value)*KPrime;
    %Param, KCisArr, KTransArr = Optimized values yielded by fminsearch.
    %Values for KCis and KTrans are independent of which parameter is
    %directly fit by fminsearch.

    Value = Value + 1;
end

KCisAve = mean(KCisArr);
KTransAve = mean(KTransArr);
KCisErr = std(KCisArr);
KTransErr = std(KTransArr);

KPrimeLigandArr = KPrime*KCisArr./KTransArr;
KPrimeLigandAve = mean(KPrimeLigandArr);
KPrimeLigandErr = std(KPrimeLigandArr);
%KPrimeLigand variables = Values for K'L.  By default, these values are
not
%set to print.  However, removing the semicolon after the above lines will
%display the KPrimeLigand variables.


%PART III: Plotting the data points
hold off;
axis([0 1 0 1]);
hold on;

DataPoints = FractionCis(:,Start:End);
for i = 1:(Points - 1)
    FractionCisMean(i) = mean(FractionCis(i,Start:End));
    Error(i) = 2*std(FractionCis(i,Start:End));
end

errorbar(FractionBound, FractionCisMean, Error, 'ko');

%Optional: This section can be decommented and the preceding section
%commented in order to toggle between displaying the binding curve as a
%scatter plot or with error bars.
%for i = 1:(Value - 1)
%    plot(FractionBound, FractionCis(:,i),'ko');
%    hold on;
%end


%PART IV: Plotting the binding curve
Fb = 0.0001:0.01:1;
%Fb = FractionBound for the fitted binding curve

Points = size(Fb);
Points = Points(2);

for i = 1:Points
    Fc(i) = ((KPrime*(KCisAve - KTransAve))/((1 + KPrime)*(KTransAve +
KCisAve*KPrime)))*Fb(i) + KPrime/(1 + KPrime);
```

```
end
%Fc = FractionCis for the fitted binding curve

KTransModel = 0;
KCisModel = Ko*(1 + KPrime)/KPrime;
%Optional: This section can be decommented and the preceding section
%commented in order to toggle between displaying a model curve displaying
%all-trans binding or all-cis binding.
%KTransModel = Ko*(1 + KPrime);
%KCisModel = 0;

for i = 1:Points
    FcModel(i) = ((KPrime*(KCisModel - KTransModel))/((1 +
KPrime)*(KTransModel + KCisModel*KPrime)))*Fb(i) + KPrime/(1 + KPrime);
end
%FcModel = FractionCis for the model in which the ligand binding is
%absolutely conformationally dependent.


plot(Fb, Fc,'k');
hold on;
plot(Fb, FcModel, '--k');
```

```
%This function returns a value corresponding to the SOS error of fitting a
%binding model to the data given using the passed value for an association
%constant in the system.  This program is based on a model in which a
%conformationally heterogenous protein binds is capable of binding to a
%non-conformationally heterogenous protein in the absence of
%self-association.  The passed parameter in this case is the association
%constant of one of the conformers.

%Param = Parameter being minimized by fminsearch (KCis)
%FractionBound = Array containing the total fraction of protein bound to
%ligand as it varies throughout the titration. (does not include the first
%point of the titration).
%FractionCis = Array containing the total fraction of protein in the cis
%conformation at it varies throughout the titration (again, does not
%include the first titration point).
%KPrime, Ko = Conformer exchange equilibrium constant and protein-ligand
%association constant, respectively (determined separately).

%Error = Sum of the squares of the difference between the value of the
%fitted curve at a titration point and the experimentally observed value.

%2003 Patrick Breheny

function Error = fitmlciserror(Param, FractionBound, FractionCis, KPrime,
Ko)

KcEst = Param;
KtEst = Ko*(1 + KPrime) - KcEst*KPrime;

Points = size(FractionBound);
Points = Points(2);

for i = 1:Points
    FractionCisPredicted(i) = ((KPrime*(KcEst - KtEst))/((1 +
KPrime)*(KtEst + KcEst*KPrime)))*FractionBound(i) + KPrime/(1 + KPrime);
end
%Generation of fitted curve based on parameters.

for i = 1:Points
   ErrorVector(i) = FractionCisPredicted(i) - FractionCis(i);
end
%Generation of error vector based on difference between fitted curve and
%real data

Error = sum(ErrorVector.^2);
%Minimization is run on the sum of the squares at each point
```